

Optimal time-activity basis selection for exponential spectral analysis: application to the solution of large dynamic emission tomographic reconstruction problems

Jonathan S. Maltz, *Member, IEEE*¹

Abstract

The clinical application of dynamic ECT reconstruction algorithms for inconsistent projection (IP) data has been beset with difficulties. These include poor scalability, numerical instability of algorithms, problems of non-uniqueness of solutions, the need to oversimplify tracer kinetics, and impractical computational burden. We present a stable, low computational cost reconstruction algorithm which is able to recover the tracer kinetics of several hundred image regions at realistic noise levels. Through optimal selection of a small set of non-negative basis functions to describe regional time-activity curves (TACs), we are able to solve for the first-order compartmental model kinetics of each region. A non-uniform resolution pixelization of image space is employed to obtain highest resolution in regions of interest. These spatial and temporal simplifications improve numerical conditioning, provide robustness against noise, and greatly decrease the computational burden of dynamic reconstruction. We apply this algorithm to IP phantom data whose source distribution, kinetics and count statistics are modeled after a clinical myocardial SPECT dataset. TACs of phantom regions are recovered to within a mean square error of 6%, an accuracy which proves sufficient to allow for quantitative detection of a myocardial perfusion defect within healthy myocardial tissue.

I. INTRODUCTION

First-order compartmental models are routinely used to describe pharmacokinetics within biological systems. The responses of such models are governed by first-order linear differential equations, usually driven by a single forcing function corresponding to the input of a particular substance into the compartmental system. As such, the responses take the form of the convolution of the sums of decaying real exponential terms with this input function. Exponential spectral analysis involves the quantification of the coefficients of these exponential terms (modes) within the compartmental model response, in terms of a preselected spectrum of exponential functions [1]. This approach obviates the need for iterative estimation of the non-linear parameters (rate constants) of a compartmental model, allowing convenient solution via linear methods. Unfortunately, owing to the fact that sums of decaying real exponentials are not uniquely parameterized in the presence of noise, solutions in terms of spectral coefficients may not be unique [2, 3]. A strong dependence may consequently exist between the particular solution obtained and the noise present within the data.

We have previously shown how unique representations of compartmental model responses may be obtained through

the representation of an exponential spectral basis using an orthogonal approximating basis [4, 5]. This technique greatly reduces problem dimension and ensures that any solution in terms of the coefficients of this basis will be unique. However, this scheme has two disadvantages:

1. The new basis functions possess negative values, and in general require negative coefficients in order to approximate an exponential spectral range.
2. The coefficients obtained do not have obvious physiological significance.

In this paper, we address the former issue by transforming the orthogonal basis set so that a broad exponential spectrum may be approximated in terms of non-negative basis (NNB) functions, linearly combined using non-negative coefficients (NNC).

We cannot address the second issue using a change of basis functions, since such a change will almost invariably lead to a representation in which the kinetic parameters will be devoid of physiological meaning. However, this consideration is arguably spurious if the time-activity curves (TACs) for all image regions are accurately recovered, since any model may then be fit to these. Our primary goal in this work is to demonstrate how the NNB-NNC representation, used in conjunction with a non-uniform resolution reconstruction grid, leads to a tremendous reduction in the dimension of the problem of direct dynamic emission computed tomography (ECT) reconstruction from inconsistent projections. By minimizing the number of parameters, this approach tends to improve problem condition, leading to an overdetermined dynamic imaging system matrix. This in turn improves robustness against noise which facilitates the recovery of TACs to an accuracy sufficient to allow clinical diagnosis based on results obtained.

II. PROBLEM FORMULATION

A. Kinetic model

Since one of the most compelling applications of algorithms for direct reconstruction from inconsistent projections (DRIP) lies in clinical myocardial ECT, we choose to address the dynamics of the first-order single compartmental model. This model is appropriate for tracers such as ^{99m}Tc-teboroxime, whose uptake and washout from the myocardium have been shown to correlate with blood flow and consequently may serve as a good indicator of myocardial defects associated with ischemia and infarcted tissue [6]. This kinetic model is represented schematically in Figure 1.

According to this model, the time-activity within region n is given by:

$$\phi_n(t) = k_1^n i(t) * e^{-k_2^n t}, \quad (1)$$

¹Jonathan S. Maltz is with the Department of Functional Imaging, Lawrence Berkeley National Laboratory, University of California, Berkeley CA 94720

where $i(t)$ is the measured blood input function, and the ‘*’ operator denotes convolution. The constants k_1 and k_2 represent the wash-in and wash-out coefficients of the compartment, respectively. We assume throughout that the blood input function has been obtained via arterial blood sampling. In cases where the projection data are reasonably consistent, the input function may be estimated from the activity within the left ventricular blood pool [7].

B. Geometric model

Without loss of generality, for purposes of illustration we refer to the 2D non-uniform resolution pixel grid shown in Figure 2.

The TAC $\phi_n[l]$ is assigned to the n th region $\Omega_n(\mathbf{x})$, $n = 1, 2, \dots, N$. In general, we require that the underlying source distribution $\Omega(\mathbf{x})$ be completely segmented into regions in this way.

C. Reconstruction problem

Given a set of projection measurements $\tilde{\mathbf{y}}(\theta(t))$ acquired at times $t = t_l$, $l = 0, 1, \dots, L - 1$, at angles $\theta[l]$, our objective is to recover the time activity curves $\phi_n[l]$ for all N regions. This amounts to reconstructing a dynamic image sequence of length L .

III. ALGORITHM FORMULATION

This algorithm represents an extension of the convolved-orthogonal basis reconstruction algorithm (COBRA) described in [5]. The formulation is very similar to COBRA, except that we now introduce a new basis and non-negativity constraints on the reconstructed TACs. We also generalize the algorithm to accommodate irregular time sampling of projections. Owing to the small number ($N < 10$) of regions present in the segmentation to which COBRA was applied in [5], these constraints proved unnecessary. However, their imposition is essential to ensure that physically realistic TACs are recovered in the solution of problems of larger dimension. For completeness, a brief overview of the COBRA framework is given, before the extensions are presented. In addition, the entire procedure for the generation of the TAC approximating basis is illustrated schematically in Figure 3.

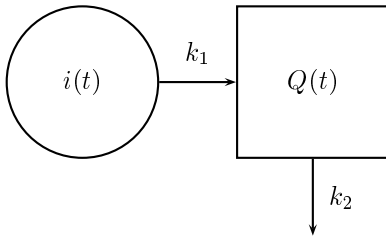


Figure 1: Compartmental model for ^{99m}Tc -teboroxime in the myocardium. Here, $Q(t)$ represents the tracer activity within the myocardium, while other symbols are defined within the text.

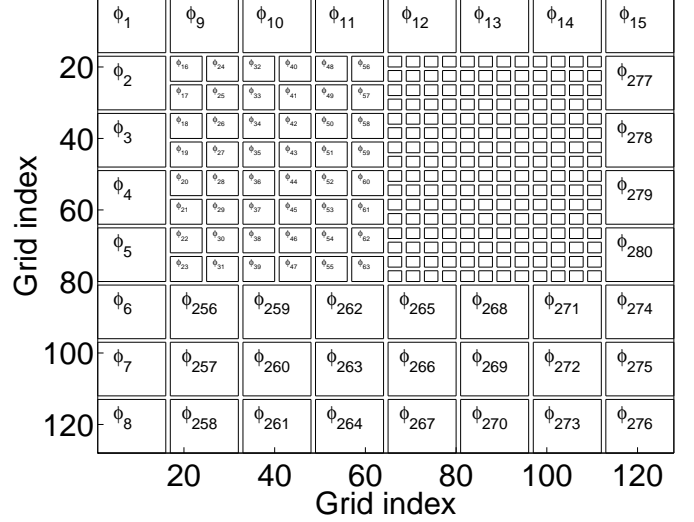


Figure 2: Non-uniform resolution pixel grid. Such a grid is proposed in order to offer increased resolution in areas of clinical interest, while reducing the overall problem dimension. Although the pixels of the highest resolution region are not labeled in the figure, region numbers are similarly assigned to thereto.

A. Spectral representation of single compartment model kinetics

Our kinetic model formulation begins by generalizing (1) to the case where the TAC of each region may be composed of linear combinations of the responses of several underlying compartments. This is often useful for modeling tissue heterogeneity, partial volume and spillover effects [8]. We thus have:

$$\phi_n(t) = \sum_{\tilde{m}=1}^{\tilde{M}} k_1^{\tilde{m}n} i(t) * e^{-k_2^{\tilde{m}} t}, \quad (2)$$

where \tilde{M} is the number of exponential modes (compartments) from which the TACs may derive.

Since the algorithm will be based in discrete time, we sample the imaging time interval at L points $t = t_l$, $t_{l+1} \geq t_l$, $l = 0, 1, \dots, L - 1$ giving:

$$\phi_n[l] = \phi_n(t) \sum_{l=0}^{L-1} \delta(t - t_l), \quad (3)$$

where $\delta(t)$ is the Dirac delta distribution.

In (2) we see that each TAC depends non-linearly on the parameter k_2 . As the projection measurements constitute sums of TAC values, they are composed of weighted sums of convolved exponential functions. The determination of the values of these parameters constitutes the extremely ill-conditioned problem of resolving the components of exponential sums [2, 3]. We wish to linearize the problem and improve its condition. To do this, we employ the exponential spectral method of Cunningham et al. and define a set of sampled exponential functions which span the range of physiologically feasible compartmental modes expected within

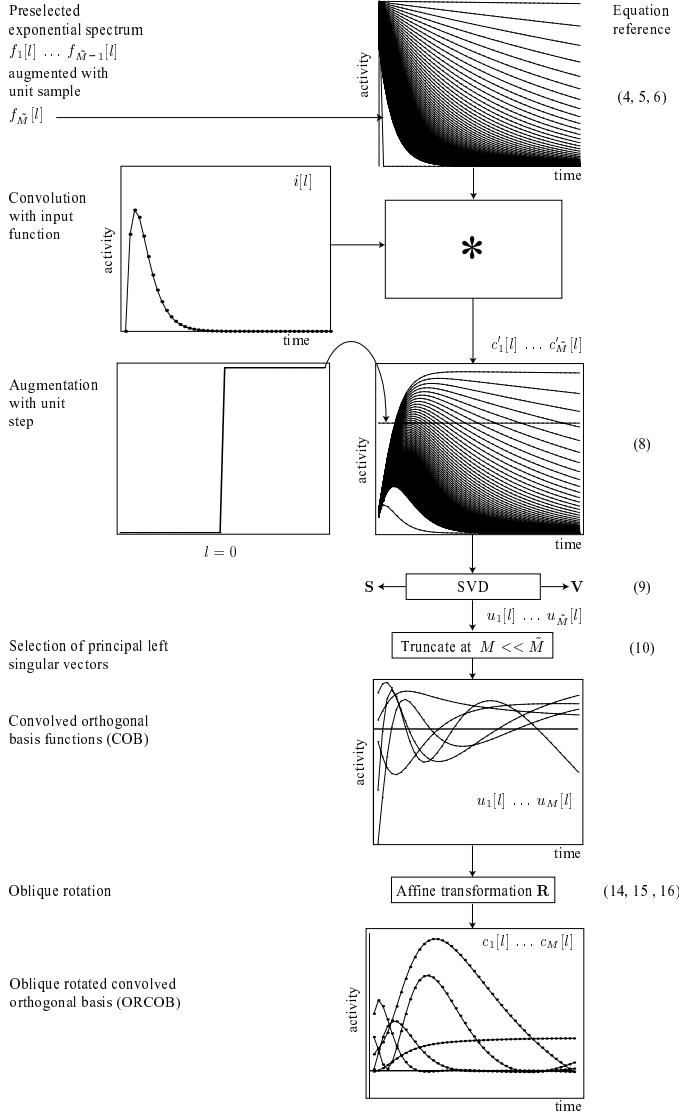


Figure 3: Schematic diagram illustrating the steps involved in generating the oblique-rotated convolved orthogonal basis. The rightmost column references the relevant equations in the text.

the imaged distribution [1]:

$$f_{\tilde{m}}[l] = e^{-k_2^{\tilde{m}} l \Delta t}, \quad l = 0, 1, \dots, \tilde{L} - 1, \quad (4)$$

$$\tilde{m} = 1, \dots, \tilde{M} - 1. \quad (5)$$

It is important that Δt be small enough to ensure adequate sampling of the spectral functions. The number of basis samples \tilde{L} must be greater than or equal to the number of projection sampling time points L , since the latter will always be a subset of the former. While the latter may be irregularly sampled, the former must be sampled regularly, for reasons that will become clear shortly.

Typically, we desire the ability to model the presence of the the blood input function within the imaged distribution. This corresponds within the spectral context to convolution with an exponential basis function (4) having $k_2^{\tilde{m}} = \infty$. This function is equivalent to the Dirac delta distribution. The spectral set (4)

is consequently augmented by:

$$f_{\tilde{M}}[l] = \delta[l], \quad (6)$$

where $\delta[l]$ is the unit sample.

We then form the $(\tilde{L} \times \tilde{M})$ matrix \mathbf{X} whose \tilde{m} th column is $f_{\tilde{m}}[l]$ as defined in (4). Convolution of each column of \mathbf{X} with the blood input function $i[l] = i(l\Delta t)$, $l = 0, 1, \dots, \tilde{L} - 1$ we form:

$$\mathbf{C}' \triangleq (\mathbf{c}'_1 \mathbf{c}'_2 \dots \mathbf{c}'_{\tilde{M}}) \quad (7)$$

where we have retained only the first \tilde{L} elements of the convolution, so that the row dimension of \mathbf{C}' is \tilde{L} .

In realistic imaging scenarios, residual activity may be present in the imaged distribution, perhaps from a previous tracer injection. In order that such an offset may be modeled, we augment \mathbf{C}' with a column:

$$\mathbf{c}'_{\tilde{M}+1} = \mathbf{1}_{\tilde{L}} \quad (8)$$

where the right-hand side is a column vector of \tilde{L} ones.

B. Orthogonalization of spectral basis

We now invoke the singular value decomposition (SVD) to find orthogonal basis vectors for the range of \mathbf{C}' . These are the left singular (column) vectors $\mathbf{u}_{\tilde{m}}$ of the SVD of \mathbf{C}' :

$$\mathbf{C}' = \tilde{\mathbf{U}} \mathbf{S} \mathbf{V}^T, \quad \tilde{\mathbf{U}} = (\mathbf{u}_1 \mathbf{u}_2 \dots \mathbf{u}_{\tilde{M}+1}) \quad (9)$$

where \mathbf{V} is the matrix of right singular vectors, and \mathbf{S} is the diagonal matrix of singular values. We associate the discrete time index l with each row of $\tilde{\mathbf{U}}$. Depending on the degree of accuracy required in the sampled representation of the $\phi_n(t)$, we utilize only the first $M \leq \tilde{M}$ of $\tilde{\mathbf{U}}$ and define:

$$\mathbf{U} \triangleq (\mathbf{u}_1 \mathbf{u}_2 \dots \mathbf{u}_M). \quad (10)$$

Typically, $M \approx 3$ to 5 is sufficient for myocardial imaging applications. Figure 4 shows the columns of \mathbf{U} when $M = 5$.

Each column of \mathbf{U} represents a basis function, and we have plotted these functions in Figure 4.

We refer to the \mathbf{u}_m as convolved-orthogonal basis functions (COBs). The COB will, in general, possess negative excursions. Since TAC values are always positive, their use in large DRIP problems requires the implementation of the constraints:

$$\hat{\phi}_n[l] = \sum_{m=1}^M \mu_{mn} u_m[l] \geq 0, \quad l \in R \quad (11)$$

where μ_{mn} is the coefficient of the m th basis function for the n th region. The value $u_m[l]$ corresponds to the (l, m) th element of \mathbf{U} . We recognize that the constraints need only be enforced at the extrema of the basis functions, as well as at the first and last sampling points. Imposition of these constraints is sufficient to preclude negative excursions. Equivalently, the set of time

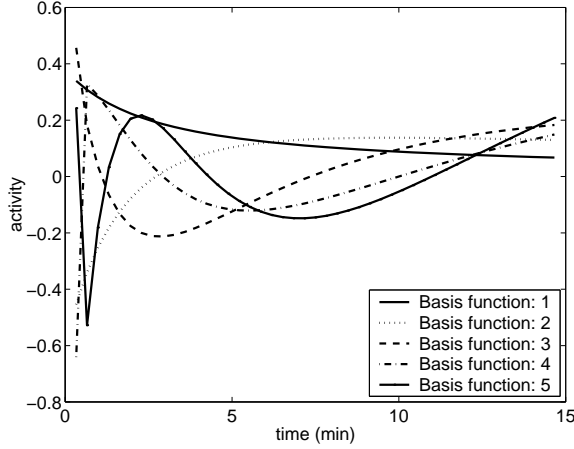


Figure 4: Many of the elements of \mathbf{U} (samples of the orthogonal basis functions) are negative, as is obvious from these plots of the columns of \mathbf{U} .

indices at which constraints are enforced is given by:

$$R \triangleq \left\{ l : \begin{array}{l} u_m[l+1] > u_m[l], u_m[l] < u_m[l-1], \\ l = 1, 2, \dots, L-2; \\ u_m[l+1] < u_m[l], u_m[l] > u_m[l-1], \\ l = 1, 2, \dots, L-2; \\ l = 0, \quad l = L-1 \end{array} \right\}.$$

The left-hand side of (11) gives the the TAC value for the n th region at discrete time index l . The value $\Xi = |R| \leq L$ corresponds to the number of time indices at which the constraints are imposed. We define the total number of constraints $\Upsilon \triangleq \Xi N$.

The constraints (11) may be rewritten as the matrix product:

$$\phi = \mathbf{A}\mu \quad (12)$$

where $\mathbf{A} \in \mathbb{R}^{\Xi N \times N}$ is a block diagonal matrix.

We desire that $\hat{\phi}_n[l]$ be non-negative for all n and $l \in R$. However, this leads to a very large matrix \mathbf{A} , whose dimension scales as the square of the number of regions. This approach consequently becomes impractical for higher resolution reconstruction problems, where the number of regions is large, unless sparse storage is used for \mathbf{A} . Even under such circumstances, the imposition of ΞN constraints is computationally burdensome.

C. Construction of non-negative basis through affine transformation

The problem described above would be greatly simplified if a non-negative basis, non-negative coefficient representation was available. To this end we begin by expressing the convolved original spectrum \mathbf{C}' as an approximation in terms of the orthogonal basis:

$$\mathbf{C}' \approx \mathbf{U} \mu_x, \quad (13)$$

where $\mu_x \in \mathbb{R}^{M \times (\tilde{M}+1)}$ is a matrix of known coefficients.

Employing the oblique rotation method proposed by Sitek et al. in [9], we introduce an invertible matrix \mathbf{R} within the identity $\mathbf{R}^{-1} \mathbf{R}$, giving:

$$\mathbf{C}' \approx \mathbf{U}(\mathbf{R}^{-1} \mathbf{R}) \mu_x \quad (14)$$

$$= (\mathbf{U} \mathbf{R}^{-1}) (\mathbf{R} \mu_x) \quad (15)$$

$$= \tilde{\mathbf{C}}(\mathbf{R}) \mu_o(\mathbf{R}). \quad (16)$$

To yield the desired basis, the elements of $\tilde{\mathbf{C}}(\mathbf{R})$ and $\mu_o(\mathbf{R})$ must obey:

$$\begin{aligned} \tilde{c}^{lm} &\geq 0, \quad l = 1, 2, \dots, \tilde{L}, \quad m = 1, 2, \dots, M \\ \mu_o^{m\tilde{m}} &\geq 0, \quad m = 1, 2, \dots, M, \quad \tilde{m} = 0, 1, \dots, \tilde{M} + 1 \end{aligned} \quad (17)$$

respectively, where \tilde{c}^{lm} is the (l, m) th element of $\tilde{\mathbf{C}}(\mathbf{R})$ and $\mu_o^{m\tilde{m}}$, the (m, \tilde{m}) th element of $\mu_o(\mathbf{R})$.

In order to find $\tilde{\mathbf{C}}(\mathbf{R})$ thus specified, we formulate the cost function:

$$\mathcal{F}(\mathbf{R}) = \max \left\{ \sum_{l=1}^{\tilde{L}} \sum_{m=1}^M (\tilde{c}^{lm})^2 \cdot I(\tilde{c}^{lm} < 0), \sum_{m=1}^M \sum_{\tilde{m}=1}^{\tilde{M}+1} (\mu_o^{m\tilde{m}})^2 \cdot I(\mu_o^{m\tilde{m}} < 0) \right\} \quad (18)$$

where $I()$ is an indicator function assuming the value unity when its argument is true. Clearly, $\mathcal{F}(\mathbf{R})$ is discontinuous, and cannot be minimized using conventional optimization algorithms. It is possible to replace $I()$ with an approximating continuous function, such as a logistic function. However, since there exist many solutions which satisfy (17), sampling algorithms which do not require cost function continuity can easily find a solution to (18). We employed Adaptive Simulated Annealing, using default algorithm parameters [10], to yield the basis shown in Figure 5. Here, ASA terminated when $\mathcal{F}(\mathbf{R})$ was less than 0.5×10^{-6} , a tolerance which proved sufficient for this application and corresponded to a maximum negative excursion of any basis function of 2.7% of the maximum value of the function. The small negative elements of $\tilde{\mathbf{C}}$ were set to zero to prevent any negative values occurring among the reconstructed TAC samples. The ASA algorithm executed in 503 minutes on a Pentium III 850MHz processor.

D. Evaluating the accuracy of TACs approximated using non-negative basis

It is important to evaluate the quality of the approximation obtained using the oblique-rotated convolved-orthogonal basis (ORCOB). To this end, we define the metric:

$$M_{\text{pow}}(z[l], \tilde{z}[l]) \triangleq \frac{\sum_{l=0}^{L-1} (z[l] - \tilde{z}[l])^2}{\sum_{l=0}^{L-1} z[l]^2} \times 100 \quad (19)$$

where $z[l]$ and $\tilde{z}[l]$ are the true and estimated functions to be compared, respectively.

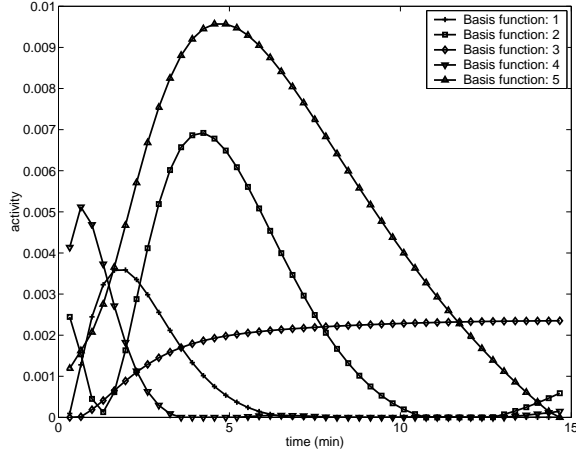


Figure 5: Oblique-rotated versions of the orthogonal basis functions which appear in Figure 4.

Let us take for example the basis set shown in Figure 5. This ORCOB was derived from the COB functions shown in Figure 4, which were in turn derived from an $(\tilde{M} - 1) = 100$ function exponential spectrum, selected by regularly sampling the interval $k_2 \in [10^{-3}, 1] \text{ min}^{-1}$. After augmentation of the spectrum with a unit sample, convolution with the input function $i[l] = l\Delta t e^{-l\Delta t/0.7}$ was performed. Subsequent augmentation with constant function (8) was followed by application of the SVD and retention of only the first five \mathbf{u}_m . This entire process is outlined in Figure 3.

We can approximate the columns of the convolved spectrum \mathbf{C}' (shown in Figure 6) using the ORCOB functions (shown in Figure 5) giving:

$$\hat{\mathbf{C}} = \tilde{\mathbf{C}}(\hat{\mathbf{R}}) \boldsymbol{\mu}_0 \quad (20)$$

where the $m\tilde{m}$ th element of $\boldsymbol{\mu}_0$ is $\mu_0^{m\tilde{m}}$ and $\hat{\mathbf{R}}$ is a solution of $\mathcal{F}(\mathbf{R}) = 0$.

Let the functions:

$$\hat{c}_{\tilde{m}}[l], \quad \tilde{m} = 0, 1, \dots, \tilde{M} + 1, \quad l = 0, 1, \dots, \tilde{L} - 1 \quad (21)$$

$$c'_{\tilde{m}}[l], \quad \tilde{m} = 0, 1, \dots, \tilde{M} + 1, \quad l = 0, 1, \dots, \tilde{L} - 1 \quad (22)$$

represent the \tilde{m} th columns of $\hat{\mathbf{C}}$ and \mathbf{C}' , respectively. We may then find the maximum approximation error among all the convolved spectral functions as:

$$M_{\text{powmax}} = \max_{\tilde{m}=0,1,\dots,\tilde{M}+1} M_{\text{pow}}(c'_{\tilde{m}}[l], \hat{c}_{\tilde{m}}[l]) = 1.3\%, \quad (23)$$

a result which is entirely satisfactory. The $c'_{\tilde{m}}[l]$ are plotted versus the $\hat{c}_{\tilde{m}}[l]$ in Figure 6.

Until this point, we have preserved the regular sampling of the basis functions to allow us to easily convolve the rotated orthogonal basis with the input function. At this juncture, we may sample the ORCOB at the projection sample time points. Element l of column vector m of the sampled ORCOB matrix \mathbf{C} is given by:

$$c_m[l] = \tilde{c}_m[k'], \quad k' = \{k : k\Delta t = t_l, k = 0, 1, \dots, \tilde{L}\} \quad (24)$$

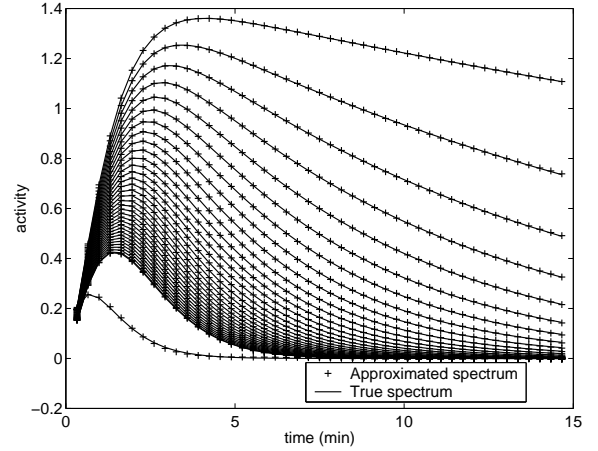


Figure 6: The approximation of the convolved spectrum using the ORCOB (+) is shown versus the original spectrum (solid). Only every third spectral function is shown for clarity.

The DRIP problem may now be reexpressed as one of estimating the coefficients μ^{mn} of the \mathbf{c}_m for all regions, so that the TAC of each region may be recovered:

$$\hat{\phi}_n[l] = \sum_{m=1}^M \hat{\mu}^{mn} c_m[l], \quad l = 0, 1, \dots, L - 1. \quad (25)$$

We must now specify how the $\hat{\mu}_{mn}$ may be estimated given the inconsistent projection measurements.

IV. PROBLEM SOLUTION

In order to solve the DRIP problem as linear system, we must construct a matrix which maps $\boldsymbol{\mu} \in \mathbb{R}^{MN \times 1}$, containing the μ^{mn} , to the projection bin measurements $\tilde{\mathbf{y}} \in \mathbb{R}^{TQ \times 1}$:

$$\tilde{\mathbf{y}} = \mathbf{F} \boldsymbol{\mu}, \quad (26)$$

where T is the total number of angular projections, and Q is the number of bins per angular projection.

The construction of $\mathbf{F} \in \mathbb{R}^{TQ \times MN}$ has been described in detail previously [5]. Briefly, this matrix serves to express each projection bin measurement in terms of the fractional contribution of each imaged region to that bin, multiplied by the fractional contribution of each basis function to each region:

$$\begin{aligned} \tilde{y}_{pq}[l] &= \sum_{n=1}^N \left[g_{pq}^n \times \sum_{m=1}^M \mu^{mn} c_m[l] \right] \\ p &= 1, 2, \dots, T \\ q &= 1, 2, \dots, Q. \end{aligned}$$

The matrix \mathbf{F} contains the geometric weighting factors g and the basis function values c arranged in such a way that the product in (26) is readily effected.

When fine pixelizations of the image space are employed, each pixel contributes to only a handful of bins within each

angular projection. Consequently, \mathbf{F} is very sparsely populated. This is a fortunate characteristic, since dense storage of this matrix becomes prohibitive for large problems.

Previous experience [5] has shown that an unweighted least squares estimator is able to provide virtually unbiased, highly efficient estimates $\hat{\boldsymbol{\mu}}$ at projection data noise levels which are typical for dynamic SPECT studies. This implies that the statistical mismatch implicit in applying the least squares estimator to Poisson data is not a significant source of estimation error.

Consequently, we choose to find the parameter estimate $\hat{\boldsymbol{\mu}}$ which solves the constrained least squares problem:

$$\theta(\hat{\boldsymbol{\mu}}) = \min_{\boldsymbol{\mu} \in Z} \|\tilde{\mathbf{y}} - \mathbf{F} \boldsymbol{\mu}\|^2 \quad (27)$$

$$Z \triangleq \left\{ \boldsymbol{\mu} \in \mathbb{R}^{MN \times 1} : \mu^{mn} \geq 0 \right\}, \quad (28)$$

where

$$\boldsymbol{\mu} = [\mu^{11}, \mu^{21}, \dots, \mu^{M1}, \mu^{12}, \dots, \mu^{MN}]. \quad (29)$$

Since all of the elements of \mathbf{F} are non-negative, the constraints (28) are sufficient to ensure that the recovered TACs in (25) are non-negative at all time points.

We henceforth refer to this algorithm for the estimation of $\boldsymbol{\mu}$ as the oblique-rotated convolved-orthogonal basis reconstruction algorithm (ORCOBRA).

V. NUMERICAL METHODS

Problem (27-28) constitutes a non-negative least squares (NNLS) problem. As such, innumerable methods exist for its solution. An excellent review of these algorithms appears in [11].

We desire an algorithm with the following characteristics:

1. Utilization of sparse matrix storage for matrix \mathbf{F} . Algorithms which do not require input of the matrix \mathbf{F} , but only the products, $\mathbf{F}\mathbf{x}$ and $\mathbf{F}^T\mathbf{y}$, are especially suited to this application. This stems from the fact that \mathbf{F} is separable into the element-by-element product of two matrices which each contain many identical blocks [12, 5].
2. Preservation of numerical precision. The algorithm should not calculate the product $\mathbf{F}^T\mathbf{F}$ [11].
3. An iterative implementation, which refines a starting estimate is desirable for warm-starting of reconstructions on higher resolution pixel grids based on coarse preliminary reconstructions obtained on lower resolution grids. We have shown previously how large computational savings ($\approx 75\%$ for 1024 pixels) may be obtained in this way [13].

We discuss four major approaches to the solution of the NNLS problem.

A. Least distance problem

The NNLS problem can be recast as a least distance problem (LDP). This is the approach taken by Lawson and Hanson [14]. In preliminary experiments using the code provided in [14], we found that the LDP-based algorithm performed poorly when applied to large sparse systems [13].

B. Quadratic program

Alternatively, problem NNLS can be reexpressed as a quadratic program (QP) [11]

$$\theta(\hat{\boldsymbol{\mu}}) = \min_{\boldsymbol{\mu} \in Q'} \boldsymbol{\mu}^T \boldsymbol{\Lambda} \boldsymbol{\mu} + \mathbf{a}^T \boldsymbol{\mu}, \quad (30)$$

with

$$Q' \triangleq \left\{ \boldsymbol{\mu} \in \mathbb{R}^{MN} : \mathbf{h} \geq \mathbf{A} \boldsymbol{\mu} \geq \mathbf{l} \right\} \quad (31)$$

where

$$\boldsymbol{\Lambda} = \mathbf{F}^T \mathbf{F}, \quad \mathbf{a} = -2\mathbf{F}^T \tilde{\mathbf{y}}; \quad (32)$$

$$\mathbf{h} = \infty, \quad \mathbf{l} = \mathbf{0}, \quad \mathbf{A} = \mathbf{I}_{MN}, \quad (33)$$

where \mathbf{I}_{MN} is an identity matrix of dimension equal to that of the parameter vector.

Since problem QP is a subproblem solved at each iteration of many constrained non-linear optimization algorithms; efficient, well-tested and numerically robust code exists for its solution. The formation of the matrix $\mathbf{F}^T \mathbf{F}$ is undesirable as it introduces numerical errors, so conventional QP methods must be modified to obviate the need to calculate this matrix. Gill et al. describe such an algorithm, which is available as part of the commercial LSSOL and NAG libraries [15, 11]. Unfortunately, this implementation is not suitable for large sparse systems.

One of the more promising QP-based algorithm tested so far is the block principal pivoting method of Portugal et al. [16], which is specifically intended for large sparse NNLS problems, and solves the linear complementary problem of the QP. In its original form, it does not have facilities for warm-starting using an initial estimate, and cannot exploit the separable nature of \mathbf{F} . Nevertheless, owing to the sparsity of $\mathbf{F}^T \mathbf{F}$, which must be stored, storage requirements are reasonable, even when ORCOBRA is applied to large datasets. Computational burden is perhaps an order of magnitude lower than that of the NAG library's dense matrix QP algorithm.

C. Richardson iteration

Among all methods investigated, the first-order Richardson method [11, p. 276] appears to offer the best compromise between speed of convergence and memory requirements. It also satisfies all of the desired characteristics specified above. The Richardson iteration is given by:

$$\hat{\boldsymbol{\mu}}^{(k)} = \hat{\boldsymbol{\mu}}^{(k-1)} + \alpha \mathbf{F}^T (\tilde{\mathbf{y}} - \mathbf{F} \hat{\boldsymbol{\mu}}^{(k-1)}) \quad (34)$$

where $\hat{\boldsymbol{\mu}}^{(k)}$ is the solution vector at iteration k and $\alpha > 0$ is a parameter affecting convergence. The iteration converges to the least squares solution under the conditions:

$$\hat{\boldsymbol{\mu}}^{(0)} \in \text{range}(\mathbf{F}^T), \quad 0 < \alpha < 2/\sigma_1^2 \quad (35)$$

where σ_1 is the largest singular value of \mathbf{F} . In a manner analogous to that proposed by Cryer [17] we obtain an NNLS solution by modifying the iteration in 34:

$$\hat{\mu}_j^{(k)} = \max \left\{ 0, \left[\hat{\mu}^{(k-1)} + \alpha \mathbf{F}^T (\hat{\mathbf{y}} - \mathbf{F} \hat{\mu}^{(k-1)}) \right]_j \right\} \quad (36)$$

$$j = 1, 2, \dots, NM$$

where $\hat{\mu}_j$ is the j th element of $\hat{\mu}$ and the operator $[\cdot]_j$ extracts the j th element of a vector argument. To avoid the difficulty of finding the largest singular value of \mathbf{F} , we initially set α to a large number and reduce it by half until a decrease in successive residuals is achieved. The initial value of α should be set high enough so that the algorithm decreases α at the first iteration. A flowchart of this algorithm appears in Figure 7.

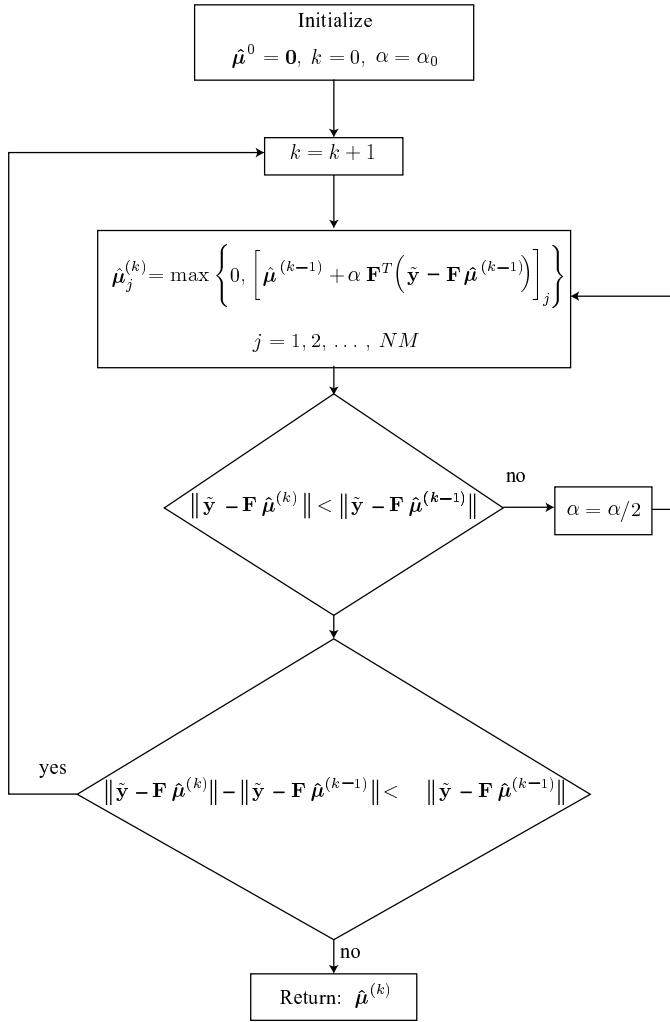


Figure 7: Flowchart illustrating the Richardson NNLS algorithm.

D. General non-linear techniques

A further alternative is to solve the NNLS problem via a general non-linear iterative optimization scheme, such as a conjugate gradient method with parameter bound constraints. An advantage of this class of methods is that it allows a more flexible selection of maximum likelihood estimation criterion.

We have not yet evaluated these methods for application to the ORCOBRA NNLS problem.

In the experiments which follow, we employ the modified first-order Richardson iteration specified in (36). We choose to store \mathbf{F} as a sparse matrix, rather than recalculating it every iteration, as this proves faster for this size dataset considering the amount of RAM memory available.

VI. ALGORITHM EVALUATION

A. Phantom imaging geometry and kinetics

We evaluate the algorithm using a phantom dataset, whose imaging and source geometries, Poisson count statistics, and tracer kinetics are modeled on an actual clinical myocardial SPECT dataset. The latter is described in detail in [7]. We choose to model our phantom after this dataset as we wish, in future work, to apply ORCOBRA to similar datasets. In so doing, we will be able to compare the results obtained to those yielded by other methods which have been used to reconstruct these data [7, 5].

A myocardial defect not present in the patient data has been included so we may determine the detectability of the defect. The regions of the phantom dataset are delineated in Figure 8, while the kinetics are shown in Figure 9.

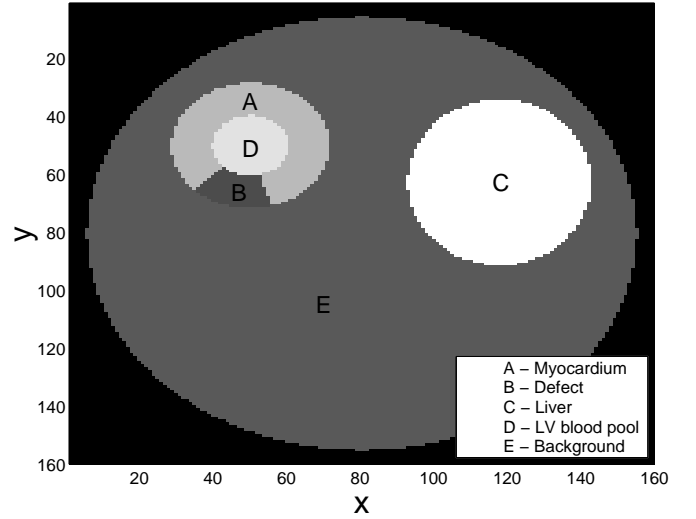


Figure 8: The phantom dataset is based on the source and imaging geometry of the clinical dynamic SPECT data.

B. Reconstruction geometry and temporal sampling

We employ ORCOBRA to reconstruct the phantom image sequence on the non-uniform resolution grids which appear in Figures 10, 11 and 12. These grids differ with respect to the maximum resolution of the sub-grid overlying the heart, the respective maximum resolutions being 40×40 , 20×20 and 16×16 pixels. A 16×16 grid overlies the liver, and an 8×8 grid, the background, in all grids.

Design of such a grid so that areas of high activity and interest such as the myocardium, left ventricular (LV) blood

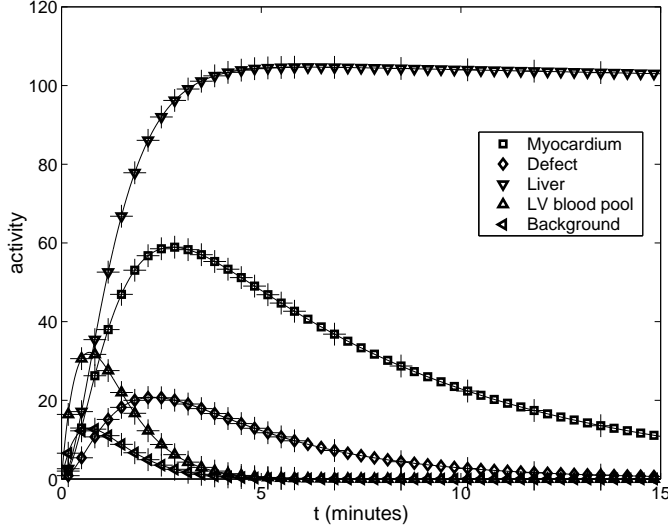


Figure 9: Phantom region kinetics appear with symbols denoting sample times. For this set, each time point corresponds to an individual camera rotation. Points marked with a '+' correspond to the sample times of those camera rotations included in the data sinogram \tilde{y} .

pool and liver receive finer discretization requires only crude localization of these areas within image space. A standard static reconstruction algorithm may be used for this purpose, or ORCOBRA may be applied to a low resolution uniform grid.

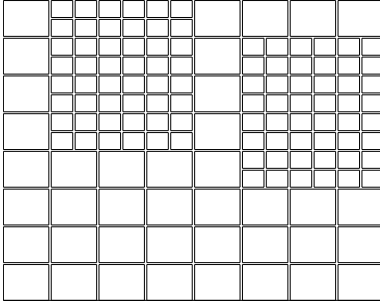


Figure 10: The coarsest grid used in these experiments offers a maximum resolution of 16×16 pixels. This grid is referred to as the 16-16-8 grid, and contains 136 pixels.

Owing to memory constraints, we include in our measured sinogram \tilde{y} , only the first 23 of the 45 single-rotation sinograms acquired by the camera. This time point selection is shown in Figure 9.

All of the 120 generated projections per rotation are utilized in the reconstruction.

A summary of the imaging parameters appears in Table 1, while parameters for the Richardson NNLS algorithm are listed in Table 2.

C. Performance metric

To evaluate goodness-of-fit between true TACs and TACs recovered by the algorithm, we base our metric on M_{pow} , which we defined in (19).

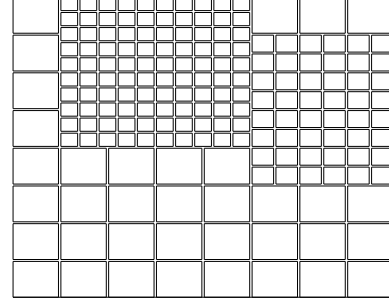


Figure 11: The 20-16-8 grid possesses a maximum resolution of 20×20 pixels and contains 184 pixels. The finest resolution region is extended in this grid so that its width spans an integral number of 8×8 -size pixels.

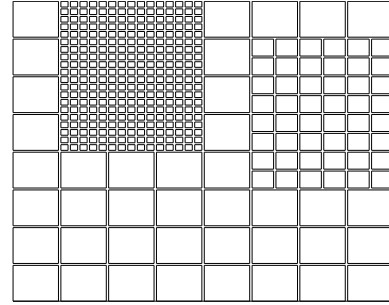


Figure 12: With a maximum resolution of 40×40 pixels, the 40-16-8 grid is the finest upon which we reconstruct the dynamic image sequence in these experiments. This grid contains 388 pixels.

Table 1
Imaging parameter values.

Parameter	Value
Imaging time (minutes)	15
Time sample points in set	45
Time sample points selected (L)	23
Rotations in set	45
Rotations selected (R)	23
Angular projections in set per rotation	120
Bins per angular projection (Q)	64

Table 2
Richardson NNLS algorithm parameter values

Parameter	Value
α_0	10
ϵ	10^{-5}

Let Ω_s , $s \in \{A, B, C, D, E\}$ represent the regions defined in Figure 8. We define the true TAC of region Ω_s as $\Phi_s[l]$, and its estimate, obtained for noise realization i by $\hat{\Phi}_s^i[l]$. We may then measure the goodness-of-fit between true and reconstructed TACs over I noise-realizations, at a particular noise level, for a particular pixel grid configuration as:

$$\bar{M}_{\text{pow}}^s = \frac{1}{I} \sum_{i=1}^I M_{\text{pow}} \left(\Phi_s[l], \frac{1}{|\Omega_s|} \sum_{n \in \Omega_s} \hat{\phi}_n[l] \right), \quad (37)$$

where $|\Omega_s|$ represents the number of pixels in region s .

D. Phantom data reconstruction results

Tables 3, 4, and 5 give values for \bar{M}_{pow}^s at various noise-levels, for the myocardium, myocardial defect and liver, respectively. For tests where noise is present, results given are averaged over $I = 30$ noise realizations. Between-trial standard deviations are also tabulated. Results for the LV blood pool and background are omitted for brevity, since these regions are usually of less interest in the diagnosis of myocardial perfusion defects.

Total count values for the simulations are given for the full 45 rotation datasets. Owing to the reduced rotation sampling scheme employed, the measured sinogram \tilde{y} to which ORCOBRA is applied contains approximately 266,000 counts. This is less than 53% of the total events recorded for the slice of patient data on which the phantom was modeled.

Figure 14 shows the mean TACs recovered for all phantom regions at this noise level. These TACs were obtained by averaging the pixel TACs within each region Ω_s in the reconstructed dynamic image sequence. Several samples of this sequence appear in Figure 13. Intraregion variability is quantified in terms of pixel TAC standard deviation from the mean, as shown in Figure 14.

Figure 15 illustrates an excerpt of the dynamic sequence obtained when ORCOBRA is applied to a single realization of a phantom sinogram. The corresponding TACs appear in Figure 16. It is clear that ‘cross-talk’ between the liver and background leads to underestimation of the liver TAC amplitude and overestimation of background activity.

All computation was performed on a Pentium III 850MHz processor equipped with 512MB of RAM and 1GB of hard disk swap space. Mean computation times for the various simulations appear in Table 6.

VII. DISCUSSION

We have shown how non-negative basis functions, which allow a spectral representation of single compartment tracer kinetics using non-negative coefficients, may be derived from a chosen exponential spectrum via the SVD and an oblique rotation. The addition to the exponential spectrum and convolved exponential spectrum of functions which allow for modeling of blood pool and regions of constant activity have not been observed to increase the dimension of the resulting

Table 3

Goodness-of-fit of the recovered normal myocardial TACs is evaluated in terms of the metric \bar{M}_{pow}^s , for various noise-levels and grid resolutions. Values are given \pm between trial standard deviations. The number of noise realizations I appears in the last column.

Counts	Pixels			Noise
	388	184	136	realizations
∞	1.37	3.72	3.11	1
5×10^9	1.90 ± 0.4	3.93 ± 0.4	3.49 ± 0.3	30

Table 4

Values of \bar{M}_{pow}^s obtained in the comparison of true and recovered myocardial defect TACs.

Counts	Pixels			Noise
	388	184	136	realizations
∞	1.46	2.86	4.80	1
5×10^9	2.74 ± 1.0	3.82 ± 1.9	5.58 ± 3.7	30

Table 5

In terms of the metric \bar{M}_{pow}^s , the TAC for the liver is the more accurately recovered than the TACs of other regions.

Counts	Pixels			Noise
	388	184	136	realizations
∞	1.26	0.69	1.23	1
5×10^9	1.41 ± 0.1	0.80 ± 0.1	1.37 ± 0.1	30

Table 6

Computational statistics derived from the experimental application of ORCOBRA to noise-free phantom data. The second column gives the time needed to calculate and store the system matrix \mathbf{F} for each grid specified in column one. Column three contains the number of Richardson iterations needed to find the solution which attains $\epsilon = 10^{-5}$, while the fourth column gives the total time to perform all iterations. The total number of floating point operations (in gigaflops) for all Richardson iterations appears in the final column.

# pixels	F prep. (min)	Iterations	Optim. time (min)	Optim. Gflops
388	16.9	309	86.1	28.3
184	4.9	236	48.9	17.5
136	3.9	310	55.2	20.2

basis set. Approximation is achieved to within a mean square error of 1.3%, and consequently constitutes a very small source of error in the dynamic reconstruction problem.

Scrutiny of both the recovered TACs in Figure 14 and the reconstructed dynamic sequence in Figure 13 reveals that the myocardial defect is easily discernible from healthy myocardium. Limited resolution, and possibly the use of the non-uniform resolution grid, leads to ‘bleeding’ of regions so that the normal myocardial TAC is underestimated while that

Table 7

Computational statistics derived from the experimental application of ORCOBRA to the 5×10^5 count phantom data. The quantities listed are described in the caption of Table 6. Where applicable, mean values over 30 tests are given \pm one standard deviation.

# pixels	F prep. (min)	Iterations	Optim. time (min)	Optim. Grops
388	16.1	107.7 ± 19.8	31.6 ± 11.0	9.9 ± 1.8
184	5.2	87.0 ± 18.1	19.9 ± 10.6	6.5 ± 1.3
136	4.2	81.2 ± 15.4	14.5 ± 2.7	5.4 ± 1.0

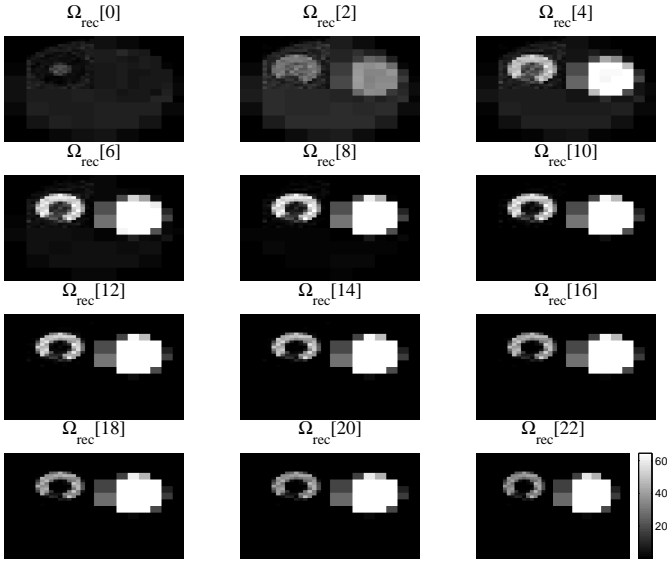


Figure 13: Snapshots of the reconstructed dynamic sequence $\Omega_{\text{rec}}[t]$, which results from the application of ORCOBRA to a dataset containing 2.66×10^5 total counts, a selected subset of a full 45 rotation, 120 projection set containing 5×10^5 counts. The source distribution was reconstructed on the 40-16-8 grid shown in Figure 12. The source intensities which appear above are the average of those obtained over 30 noise realizations. The myocardial defect is well differentiated from the surrounding healthy tissue.

of the less intense LV blood pool is overestimated. Similarly, the liver TAC is underestimated, and its activity clearly bleeds into the surrounding background, leading to overestimates of the activity in the latter. However, mean square errors for these clinically relevant regions are below 6%, even though only half the realistic number of photon counts for a single slice $^{99\text{m}}\text{Tc}$ -teboroxime myocardial study are utilized. Figures 15 and 16 show results analogous to those which appear in Figures 13 and 14, for the case where ORCOBRA is applied to a single noise realization.

Table 7 indicates that the highest resolution 40-16-8 grid, which is illustrated in Figure 12, is reconstructed in a mean time of 48 minutes on a mid-speed Pentium III CPU. This includes the 16 minutes needed to calculate and store \mathbf{F} , which may then be used in multiple subsequent applications of the algorithm to datasets for which the projection geometry and

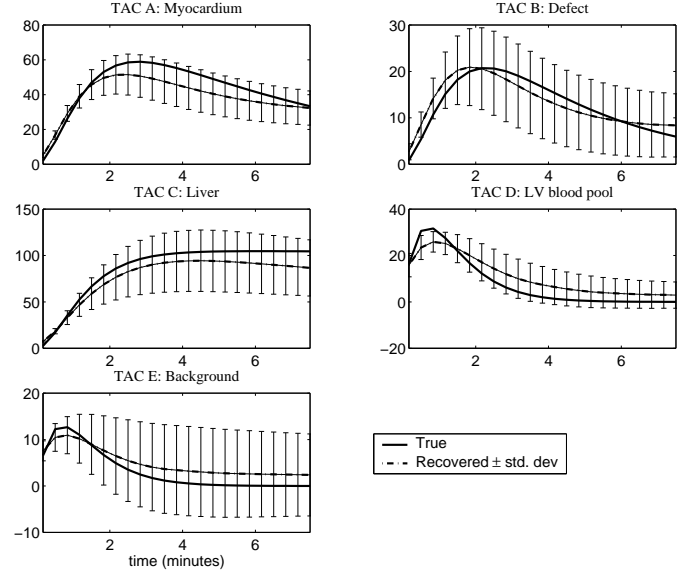


Figure 14: TACs obtained by finding the mean TAC of each region of the full dynamic reconstructed sequence, several time samples of which appear in Figure 13. Standard deviations plotted relate to intraregion variation of individual pixel TACs. Although ‘bleeding’ of regions leads to the overestimation of the activity in less intense regions at the expense of underestimating regions of high activity such as the healthy myocardium and liver, recovered TAC quality is sufficient to allow isolation of the myocardial defect.

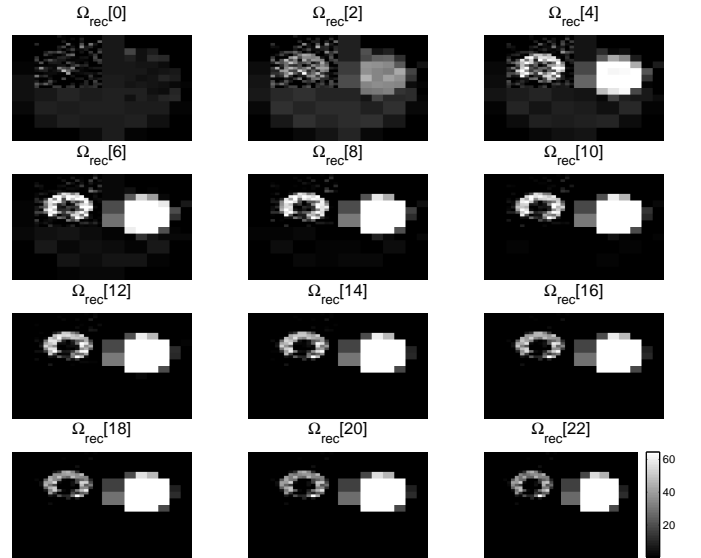


Figure 15: Snapshots of the reconstructed dynamic sequence $\Omega_{\text{rec}}[t]$, which results from the application of ORCOBRA to a single sinogram containing 2.66×10^5 total counts. Corresponding TACs are shown in Figure 16.

general kinetic model are appropriate. The chosen Richardson termination tolerance of $\epsilon = 10^{-5}$ proved overly stringent and execution times may be halved without materially affecting the quality of the reconstruction by selecting a more reasonable

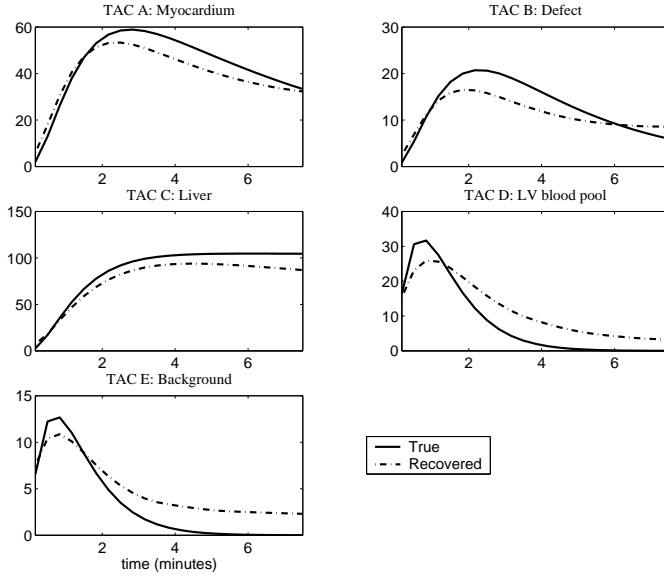


Figure 16: Mean TACs for each region recovered from a single sinogram realization at 2.66×10^5 total counts.

value for this parameter, such as $\epsilon = 10^{-3}$. Comparing Table 6 (application to noisy-free data) and Table 7 (application to noisy data) we see that the number of iterations required to reach the termination criterion in the noise-free case is greater. This is probably due to the greater smoothness and convexity of the cost function surface in the noise-free case, which allows a better solution to be found, albeit after a longer search process. For datasets of this size, computation times are likely to be significantly improved by increasing the available RAM memory to 1.5GB, as this will eliminate the delays due to memory swapping between RAM and hard disk swap space experienced here.

The 503 minutes required to find the oblique-rotated basis is prohibitive in clinical applications, especially since the ORCOB is dependent on the blood input function and must be recalculated for each study. This problem may be overcome by applying the SVD to the original exponential spectrum (4) rather than the convolved spectrum. An affine transformation may then be found that produces a non-negative basis set in exactly the same manner described in Section III-C. This set is appropriate for a wide range of studies, as it may be convolved with the input function for a particular study to form a TAC basis. Such a solution for the exponential spectrum chosen in Section III-D was found in 42 minutes, and had similar approximating performance to the ORCOB used in the simulations. However, modeling of constant TAC offset function (8) is understandably poor, since this function cannot be modeled in the unconvolved spectrum without knowledge of the input function. Accurate approximation of such functions requires augmentation of the ORCOB with a constant function. This has the disadvantage of increasing the size of the inverse problem by a factor of $(M + 1)/M$.

A second characteristic of the algorithm which detracts from its clinical utility is the requirement that a good estimate

of the blood input function be available. Accurate measurement of this function via arterial blood sampling is logistically difficult and significantly increases the cost and complexity of ECT studies. An alternative is to estimate the blood input function from the activity within the left ventricular blood pool. This is difficult in the case of the DRIP problem, since the activity within this region cannot be obtained from an image series. Reutter et al. have shown how a basis of 16 B-splines are able to approximate with reasonable accuracy the regional TACs of myocardial studies [12]. While general approximating functions such as B-splines do not enforce consistency with a particular kinetic model, they appear to be adequate for the solution of the DRIP problem when the number of regional TACs to be estimated (and hence the size of the inverse problem) is small. In cases where the application of standard (static) reconstruction algorithms to inconsistent projection data is sufficient to allow the image domain to be crudely segmented into bulk regions (myocardium, liver, LV blood pool and background), reasonable estimates for the blood input function might be obtained directly from projections using the method of Reutter et al. ORCOBRA could then be invoked to produce higher resolution reconstructions directly from the projection data. The consistency between the mean TAC of the LV blood pool obtained by ORCOBRA, and the input function estimate might serve as a measure of the accuracy of the latter.

Future work will involve the development of second-order iterative methods in order to increase the convergence rate of ORCOBRA. The algorithm will also be further evaluated through its application to the clinical data upon which the phantom was modeled.

VIII. ACKNOWLEDGMENTS

The invaluable assistance of Drs. BW Reutter and RH Huesman in the procurement of data is greatly appreciated, as is the support and encouragement of Dr. TF Budinger, who motivated this research and kindly assisted in the preparation of this paper. I would also like to thank Dr. GT Gullberg and Dr. EVR Di Bella of the Department of Radiology at the University of Utah for providing the patient data on which the phantom is based, and Mr. MR Darmalingum for assisting in the preparation of illustrations for this paper.

This work was supported in part by the National Heart, Lung, and Blood Institute of the U.S. Department of Health and Human Services under grants HL-07367, R01-HL50663 and P01-HL25840 and in part by the Director, Office of Science, Office of Biological and Environmental Research, Medical Sciences Division of the U.S. Department of Energy under contract DE-AC03-76SF00098.

IX. DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government and the National Institutes of Health. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the National Institutes of Health, nor

The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, or The Regents of the University of California.

Ernest Orlando Lawrence Berkeley National Laboratory is an equal opportunity employer.

X. REFERENCES

- [1] V. J. Cunningham and T. Jones, "Spectral analysis of dynamic PET studies," *Journal of Cerebral Blood Flow and Metabolism*, vol. 13, pp. 15–23, 1993.
- [2] C. Lanczos, *Applied analysis*. Prentice-Hall, 1956.
- [3] J. G. Reich, "On parameter redundancy in curve fitting of kinetic data," in *Kinetic data analysis : design and analysis of enzyme and pharmacokinetic experiments* (L. Endrenyi, ed.), ch. IV, pp. 39–60, Plenum Press, 1981.
- [4] J. S. Maltz, E. Polak, and T. F. Budinger, "Multistart optimization algorithm for joint spatial and kinetic parameter estimation from dynamic ECT projection data," in *IEEE Nuclear Science Symposium and Medical Imaging Conference Record*, vol. 3, pp. 1567–1573, 1998.
- [5] J. S. Maltz, "Direct recovery of regional tracer kinetics from temporally inconsistent dynamic ECT projections using dimension-reduced time-activity basis," *Physics in Medicine and Biology*, vol. 45, pp. 3413–3429, November 2000.
- [6] A. M. Smith, G. T. Gullberg, P. E. Christian, and F. L. Datz, "Kinetic modeling of teboroxime using dynamic SPECT imaging of a canine model," *Journal of Nuclear Medicine*, vol. 35, pp. 484–495, Mar. 1994.
- [7] B. W. Reutter, G. T. Gullberg, and R. H. Huesman, "Kinetic parameter estimation from dynamic cardiac patient SPECT projection measurements," in *1998 IEEE Nuclear Science Symposium and Medical Imaging Conference Record*, pp. 1953–1958, 1999.
- [8] F. O'Sullivan, "Imaging radiotracer model parameters in PET: A mixture analysis approach," *IEEE Transactions on Medical Imaging*, vol. 12, pp. 399–411, Sept. 1993.
- [9] A. Sitek, E. V. R. D. Bella, and G. T. Gullberg, "Factor analysis of dynamic structures in dynamic SPECT imaging using maximum entropy," in *IEEE Nuclear Science Symposium and Medical Imaging Conference*, pp. 2227–32, 1999.
- [10] A. L. Ingber, "Very fast simulated re-annealing," *Journal of Mathematical Computer Modelling*, vol. 12, no. 8, pp. 967–973, 1989.
- [11] A. Björck, *Numerical methods for least squares problems*. SIAM, 1996.
- [12] B. W. Reutter, G. T. Gullberg, and R. H. Huesman, "Direct least squares estimation of spatiotemporal distributions from dynamic cardiac SPECT projections using spatial segmentation and temporal b-splines," *IEEE Transactions on Medical Imaging*, vol. 19, no. 5, pp. 434–50, 2000.
- [13] J. S. Maltz, "Multiresolution constrained least-squares algorithm for direct estimation of time activity curves from dynamic ECT projection data," in *Proceedings of the SPIE - The International Society for Optical Engineering, Medical Imaging 2000: Image Processing*, vol. 3979(24), pp. 586–598, 2000.
- [14] C. L. Lawson and R. J. Hanson, *Solving least squares problems*. Prentice-Hall, 1974.
- [15] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright, "Procedures for optimization problems with a mixture of bounds and general linear constraints," *ACM Transactions on Mathematical Software*, vol. 10, pp. 282–298, 1984.
- [16] L. F. Portugal, J. J. Júdice, and L. N. Vicente, "A comparison of block pivoting and interior-point algorithms for linear least squares problems with non-negative variables," *Mathematics of Computation*, vol. 63, no. 208, pp. 625–643, 1994.
- [17] C. Cryer, "The solution of a quadratic programming problem using systematic overrelaxation," *SIAM Journal of Control and Optimization*, vol. 9, pp. 385–92, 1971.